Digital research infrastructure of two research laboratories in the field of structural dynamics

Alain Batailly[1,*], Mathias Legrand[2], Solène Kojtych[1], Benoît Charbonneau[1]

Abstract
This white paper presents the Digital Research Infrastructure (DRI) of two research laboratories in the field of structural mechanics: the Laboratory for Acoustics and Vibration Analysis at Polytechnique Montréal and the Structural Dynamics and Vibration Laboratory at McGill University. Both laboratories face similar challenges as their research activities are centered on the numerical simulation of complex industrial engineering systems. While the presented DRIs are key for the development of best practices in research data management and the diffusion of research tools and articles, their sustainability is threatened by the lack of funding programs to promote such initiative. This paper briefly describes the current state of the DRI in these labs, what is envisioned for the future and what are the main challenges ahead.

1 - Département de génie mécanique, Polytechnique Montréal, P.O. Box 6079, Succ. Centre-Ville, Montréal, Québec, Canada H3C 3A7
2 - Département de génie mécanique, Université McGill, 817 rue Sherbrooke ouest, Montréal QC H3A 0C3
∗ contact author: Alain Batailly, alain.batailly@polymtl.ca

Current DRI at the Laboratory for Acoustics and Vibration and the Structural Dynamics and Vibration Laboratory

Context

The Laboratory for Acoustics and Vibration Analysis (LAVA) at Polytechnique Montréal is a university research laboratory that, in part[1], focuses on the numerical prediction of the vibration behaviour of sophisticated nonlinear mechanical systems. This computationally intensive research is carried out in close collaboration with several industrial and academic partners in North America and Europe. The LAVA currently employs one research associate, two postdoctoral fellows, five Ph.D. candidates and two Master's students. Other undergraduate, graduate, domestic and/or international students regularly join our group for short term stages or research internships all year long.

The Structural Dynamics and Vibration Laboratory (SDVL) at McGill University is also a university research laboratory concerned with theoretical knowledge as well as the development of new tools and methods capable of bringing solutions to practical problems in the industrial sphere. Its research activities include vibration, acoustics, structural dynamics, nonlinear dynamics, and wave propagation. Its research team is mostly composed of Ph.D. students, postdoctoral fellows and undergraduate students conducting research internships. Both laboratories rely on a similar DRI that is hereafter referred to as the LAVA DRI for the sake of brevity.

Specific DRI needs

The numerical research carried out at LAVA calls for a specific work environment in order to accommodate the needs of each lab member as well as academic and industrial collaborators. In particular, the handling of proprietary and confidential industrial data imposes stringent confidentiality requirements. This becomes a key issue when a research project involves export control data. These confidentiality requirements, along with the other specific needs listed below—*training*, *accessibility*, *sustainability* and *adaptability*, have shaped the LAVA DRI.

(1) LAVA's primary mission is to *train* highly qualified personnel (HQP) in the field of numerical structural dynamics. Its DRI thus promotes an environment centered on the use of open-source numerical tools—programming languages and software packages—in order to maximize students and researchers' autonomy in their current work environment (*e.g.*, within university) and future work environments where commercial software packages may or may not be readily available. The use of open-source numerical toolboxes could also alleviate roadblocks related to proprietary data formats and specific license restrictions while giving lab members the opportunity to contribute to the development of these tools. (2) Given the highly collaborative nature of our research program at LAVA, all lab members and collaborators must also be able to easily and securely *access* their data, both locally and remotely.

[1]More specifically, the presented DRI is used by one of the two research teams of the LAVA, this research team is referred to as LAVA for the sake of simplicity in this paper. The second research teams only uses selected components of the DRI such as the wiki website.

In order to fully benefit from remote access, without critically depending on a network connection bandwidth, specific tools and methods must be given to the lab members. (3) The issue of *sustainability* is critical in any research environment. The LAVA DRI has been conceived in a way to facilitate the exchange of data between lab members and promote collaborative work. For large-scale industrial projects that may involve several researchers alongside graduate and undergraduate students, there is a need for robust development and archival procedures for data transmission to new collaborators. Finally, (4) one key aspect of any DRI in an academic context is its *adaptability*. While students and researchers must be provided with the basic tools to complete their research project, it is essential that they are given the ability to leverage the DRI to incorporate new tools and methods to innovate for the benefit of other lab members. Lab members must be able to easily install new software packages and have access to efficient Integrated Development Environments (IDE) for a variety of programming languages.

The LAVA DRI

The LAVA DRI has been set up with the invaluable assistance from the Information Technology (IT) services of the mechanical engineering department at Polytechnique Montréal in 2016. IT services play a key role in the definition of hardware requirements as well as the ongoing maintenance of the different computers and servers.

Overall architecture

The LAVA DRI is schematically represented from a user perspective in Fig. 1. It involves three key actors: lab members, the general public and industrial collaborators. It is centered on a dedicated 120 TB data server, two computing servers (for a total of 176 cores) and fifteen workstations featuring 12 to 16 cores with 16 to 64 GB of RAM memory. All computers and servers run on Linux-based operating systems, which greatly facilitates sharing computational resources between lab members as well as remote access. Authentication on all the workstations and servers is ensured by a Network Information Service (NIS) server, not represented in Fig. 1. User data is stored on the data server so that any lab member may access relevant data from any workstation or from anywhere through ssh.

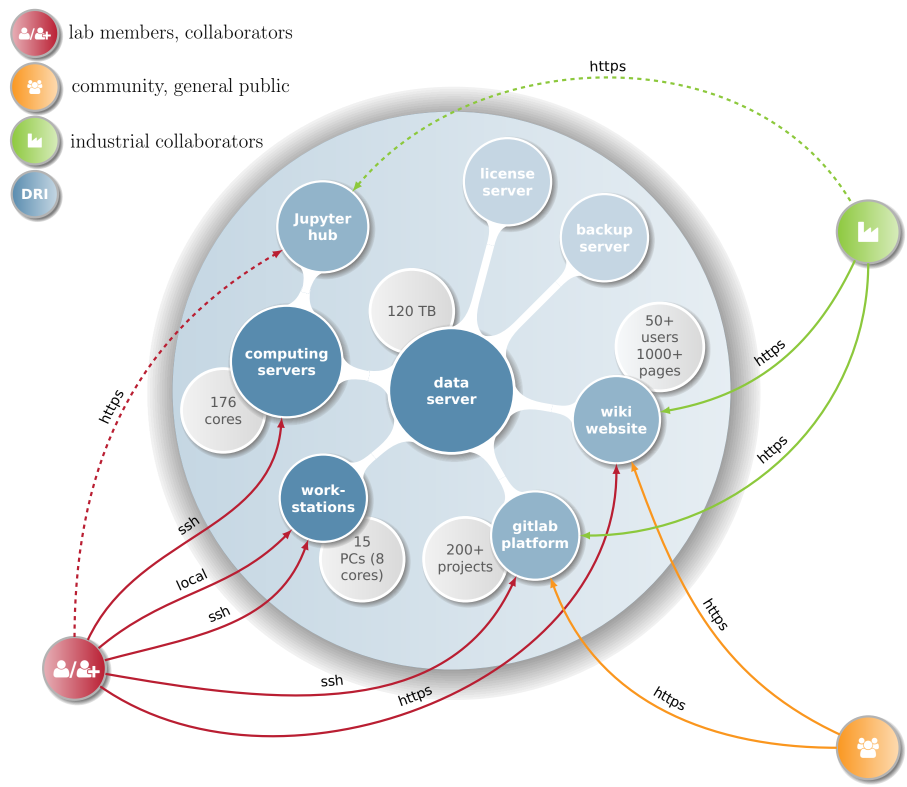

Figure 1. LAVA's DRI: users' perspective (Jupyter hub is not yet implemented)

At the heart of the LAVA DRI lie a wiki website[2] and a Gitlab platform[3]. (1) The wiki website is an essential tool for the development of good research practices and communication, both internally and externally. Specific secured pages can be easily created for a given project so that the wiki website allows to communicate efficiently with industrial and academic partners on the progress of research activities. Within the wiki website, each member possesses personal space that may be used for presentation purposes, progress archival, reports or simply as an electronic notebook. It is a go-to place for the lab members who can find custom-made tutorials and resources for the lab tools. It is used on a daily basis by more than 50 collaborators worldwide. (2) Because most of the lab research

[2]https://lava-wiki.meca.polymtl.ca/, which runs on the dokuwiki wiki software.
[3]https://gitlab.lava.polymtl.ca

activities are, in a way or another, collaborative (co-authored LaTeX documents, shared development of codes...), and version control is now a standard for the development of any numerical codes, the LAVA Gitlab platform is central for most lab members who are working on multiple projects simultaneously. It is accessible from anywhere in the world and currently hosts more than 200 projects. The Gitlab platform and the wiki website have been pivotal tools for all research activities during the Covid-19 pandemic. The wiki website and the Gitlab platform are the LAVA interface with the general public and industrial collaborators. The last two components of the LAVA DRI are departmental infrastructure including a backup data server and a license server for the use of commercial software packages.

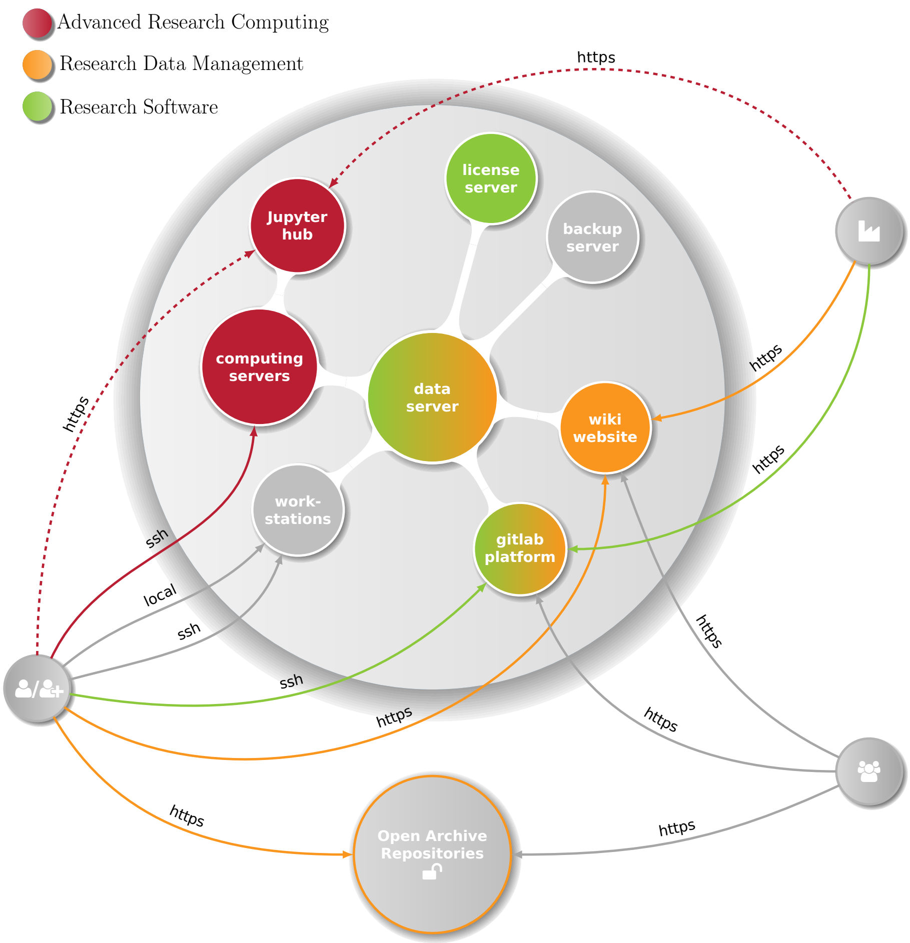

Figure 2. LAVA's DRI: functional overview

Through to the LAVA DRI, the impact of the Covid-19 pandemic on our ongoing research activities has been minimal. The remote access to all data, computers and servers was fully operational prior to the pandemic. Students at home with a regular internet connection are able to access their workstation through the ssh protocol with very low CPU load on a standard laptop. In this context, two actions were taken to minimize the impact of remote work for students: the purchase of additional monitors and the set-up of an internal and secured real-time messaging application[4] to optimize communications among lab members.

Detailed description

The LAVA DRI—which was mostly funded thanks to industrial support—was set up a few months after the Canada Foundation for Innovation (CFI) released its perspective for the development of a DRI strategy for Canada[5]. Because CFI had a vision for a Canada-wide infrastructure, not all of the identified seven key components of a DRI are relevant at the scale of a single research laboratory. The focus was made on four of those components: (1) providing lab members with *Advanced Research Computing* (ARC) capacities, (2) promoting best practices in *Research Data Management* (RDM) and dissemination of (3) *Research Software* (RS) while keeping the focus on (4) the *training of HQP*. A conceptual representation of the LAVA DRI related to these four elements is depicted in Fig. 2. Since most of the laboratories research activities relate to the development of efficient numerical tools for complex engineering applications, ARC is essential to collaborate with the industry. Contrary to numerical simulations in computational fluid dynamics for instance, typical numerical simulations run at LAVA require raw CPU power more than a parallel architecture. This motivates the use of personal computing servers that may also be used remotely by industrial collaborators for the test of numerical tools prior to their integration within their own environment. To this end, a Jupyter-hub server—yet to be implemented—will complement LAVA's ARC by providing collaborators an easy way to run their codes remotely.

[4]This messaging application is based on https://rocket.chat/.
[5]https://www.innovation.ca/sites/default/files/Funds/cyber/developing-dri-strategy-canada-en.pdf

Beside of industrial collaborations, a significant part of the research carried out at LAVA leads to the development of numerical tools and research software that are shared between lab members, collaborators and even the general public when the codes have reached a certain degree of maturity. During their development, codes are accessible to certain lab members on the Gitlab platform. Related tutorials or case studies are stored in the wiki website. While repositories of the Gitlab platform may be shared with the general public, it is usually preferred to disseminate a first release version of the codes through external open archive repositories compatible with version control[6]. This advantageously allows to obtain a sustainable permanent identifier for the codes.

Finally, and most importantly, the LAVA DRI promotes best practices in RDM. It is indeed in agreement with key aspects of the *FAIR*[7] principle. Findability and Accessibility of data are both guaranteed by version control of all the lab's projects as well as the extensive use of the wiki website, while carefully safeguarding confidential and proprietary data. This notably prevents loss of knowledge when a student or a researcher leaves the lab. The choice of programming language and the promotion of open-source software packages are essential to ensure Interoperability and Reusability. All published articles, student theses, as well as non-confidential technical reports of the LAVA are deposited in open archive repositories in compliance with the Tri-agency open access policy[8].

Challenges and roadblocks

One of the biggest challenges related to the LAVA DRI is its cost and maintenance, both in time and money. Financially, the initial investment was only made possible thanks to a significant financial support from industrial partners since there was no available public sources of funding for setting up such infrastructure. Over the past five years, the LAVA DRI represented an investment of about $\$25,000$ CAD every year due to the purchase of three servers. Nonetheless, it is worth mentioning that this remains a very competitive investment for research labs for which university IT services may bill up to $330 per Terabyte of data per year, for data storage only.

Keeping the DRI fully functional for lab members is also extremely time-consuming. While there is a significant support from IT services from the mechanical engineering department for the heaviest tasks (server configuration and initial set-up, maintenance operations and disk replacements...), most of the daily maintenance (software updates and installations, updates of the operating systems every six months or so, set up of custom-made programming environments with optimized compilation procedures, building and maintenance of the wiki website...) is carried out by the professor leading the research team. The lack of funding programs to hire a dedicated DRI coordinator (it is estimated that the LAVA DRI would require half a full-time employee) or for supporting these research infrastructures implies:

1. a very significant *workload on professors* (up to 10 % of their working hours may be dedicated to the DRI, particularly when implementing new tools and platforms),
2. a *financial burden on the research labs* that must dedicate a significant amount of their budget for the maintenance of the DRI and its development,
3. an overall *roadblock for innovation* due to limited resources.

Vision for the future

Shared resources at university, provincial or even federal levels are beneficial as they may lower the cost of certain types of infrastructure (such as large parallel computing facilities) but should not be constraining for the training of HQP. In that sense, the possibility for research labs to rely on their own DRI, and their own development environment, is essential to promote innovation and the development of new tools and technologies. A diversity of efficient DRI, reflecting the distinct context and practices in each scientific domain, will help the training of a new generation of researchers with extensive knowledge in RDM and ARC.

The promotion of best practices in RDM and RS truly implies a paradigm shift (1) to make all published articles fully accessible to the general public on central public repositories, (2) to promote the use of open-source software packages, (3) to disseminate in a sustainable manner open-source codes and research software, and (4) to ensure the continuity of research activities when a researcher or a student leaves a lab. While significant progress has been made in terms of making research articles freely accessible online, Canadian researchers are still missing a central repository to deposit their articles and receive a sustainable and permanent identifier[9]. There is also a lack

[6] Such as the international platform https://www.softwareheritage.org/
[7] Findability, Accessibility, Interoperability, Reuse
[8] https://www.ic.gc.ca/eic/site/063.nsf/eng/h_F6765465.html
[9] For many researchers, the French HAL repository https://hal.archives-ouvertes.fr is a reference in this area.

of incentive for researchers to disseminate their codes and software packages.

How to bridge the gap

As both LAVA's and SDVL's research groups are welcoming an increasing number of students and researchers, developing new functionalities and making their DRIs sustainable will require manpower. In addition to the required technical skills (IT), considering the challenges ahead in terms of RS and data archival, the hired person—that would ideally be working for both the LAVA and the SDVL—will also have to be an expert in the fields of RDM and scientific computing.

In order to maximize the efficiency of the LAVA DRI, it is planned to separate the data server from a new application server that would run: the wiki website, the Gitlab platform, all virtual machines and servers, the instant messaging platform and the Jupyter hub server. This will be a major improvement that would offer a greater security and flexibility in terms of implementation of new functionalities: an automated backup of the virtual servers would notably prevent any loss of data or crash on the Gitlab platform and the wiki website. It would advantageously yield an even more secure environment by compartmentalizing the distinct entities of the DRI. However, this application server represents an investment of about $\$45,000$ CAD for which the LAVA will need to request specific funding to a yet-to-be-identified funding agency.

Other improvements of the DRI are expected, specifically in terms of authentication procedures. Connecting the LAVA DRI to a central Active Directory server would allow a unified management of user accounts for all components of the DRI (workstation logins, wiki website, Gitlab platform...) and thus significantly facilitate the registration of each new user or collaborator.

In terms of software, because both the LAVA and the SDVL intend to develop open-source tools and methodologies in their research fields, it is expected to reduce to the strict minimum the dependence on commercial software packages. Such software will only be used when requested by an industrial partner. This will lead to significant savings every year and an optimal distribution of research funds. The LAVA DRI will keep using a versatile development environment relying on the use of the Python programming language, with an optimized compilation of scientific libraries (by means of Intel Python MKL[10]) and custom environments (implementing Conda[11]). In the short term, the implementation of the Jupyter hub server will bridge the lab research tools to industrial environments allowing engineers to run codes directly on LAVA's server, thus preventing potentially time-consuming support for the developed codes that are run within an industrial environment.

[10] https://software.intel.com/content/www/us/en/develop/blogs/python-optimized.html
[11] https://docs.conda.io/en/latest/